Facilitating efficient Mars terrain image classification with fuzzy-rough feature selection

Barnes, David Preston; Shang, Changjing; Shen, Qiang

Published in:
International Journal of Hybrid Intelligent Systems

DOI:
10.3233/HIS-2011-0126

Publication date:
2011

Citation for published version (APA):
Barnes, D. P., Shang, C., & Shen, Q. (2011). Facilitating efficient Mars terrain image classification with fuzzy-rough feature selection. *International Journal of Hybrid Intelligent Systems*, *8*(1), 3-13. https://doi.org/10.3233/HIS-2011-0126

Facilitating Efficient Mars Terrain Image Classification with Fuzzy-Rough Feature Selection

Changjing Shang*, Dave Barnes and Qiang Shen

Dept. of Computer Science, Aberystwyth University, Wales, UK.
{cns,dpb,qqs}@aber.ac.uk

Abstract. This paper presents an application study of exploiting fuzzy-rough feature selection (FRFS) techniques in aid of efficient and accurate Mars terrain image classification. The employment of FRFS allows the induction of low-dimensionality feature sets from sample descriptions of feature vectors of a much higher dimensionality. Supported with comparative studies, the work demonstrates that FRFS helps to enhance both the effectiveness and the efficiency of conventional classification systems such as multi-layer perceptrons and K-nearest neighbors, by minimizing redundant and noisy features. This is of particular significance for on-board image classification in future Mars rover missions.

Keywords: Mars images; image classification; fuzzy-rough feature selection

1 Introduction

Over the last decade, significant insights have been gained into the potential for past or present Martian life and the capabilities of the Mars environment to sustain a long-term human or robotic colonized presence. A substantial amount of such information comes from images obtained by the front-line Panoramic Camera (Pancam) instruments which serve as the scientist's eyes on Mars [2, 12]. Successful exploitation and application of these images requires a high degree of accuracy and precision in their analysis. This is both time consuming and prone to human errors. Automated analysis of such images has thus become an important task, especially for surveying places (e.g. for geologic cues) in Mars [17, 31].

A key element of analyzing Pancam terrain images is to detect rocks and other objects captured in such images. Correct analysis of these images provides useful semantic descriptions of the physical nature of the terrains. This helps to uncover knowledge of the surrounding terrain of Martian rovers which will in turn, help to exploit their mobility capabilities. However, rocks and other matters on Mars exhibit diverse morphologies, colors and textures. They are often covered in dust, grouped into self-occluding piles or partially embedded in

* Corresponding author

terrain. Also, Mars terrain images vary significantly in terms of intensity, scale and rotation, and are blurred with measurement and transmission noise. These factors make Martian image classification a challenging problem. Any progress towards automated detection and recognition of image regions which would correspond to whole or parts of Martian rocks, and their types and surroundings, will make a significant contribution to the accomplishment of addressing this challenge [32].

Many approaches may be applied for image classification. However, it is difficult to predict which particular set up and what techniques would be most effective for large-scale Mars images. Therefore, it is useful to build different classifiers and to validate their performance with respect to common criteria such that a likely optimal mechanism can be identified. For this purpose, application-oriented studies have been carried out to investigate and compare the use of two types of popular and useful classifier: Multi-layer neural networks and K-nearest neighbors. Note that these well-developed image classification methods are intentionally employed here, in order to reduce potential mission risk. This is due to the observation that flight projects normally opt to use existing mature technologies rather than totally new mechanisms which tend to have limited experimental performance data.

One critical step to successfully build an image classifier is to extract and use informative features from given images [11, 16, 18]. To capture the essential characteristics of such images, many features may have to be extracted without explicit prior knowledge of what properties might best represent the underlying scene reflected by the original image. However, generating more features increases computational complexity, especially in light of on-board processing of Mars images where demand for computational memory and processing time must be minimized, despite the nowadays generally available and relatively cheap computer power. Besides, not all such features may be useful to perform classification [13]. Due to measurement noise the use of extra features may even reduce the overall representational potential of the feature set and hence, the classification accuracy. Thus, it is often necessary to employ a method that can determine the most significant features, based on sample measurements, to simplify the classification process, while ensuring high classification performance.

This paper presents an integrated approach for performing large-scale Mars terrain image classification, by exploiting the recent advances in fuzzy-rough feature selection techniques [15]. It is based on the initial work presented in [28, 29], where fuzzy-rough methods were, for the first time, applied to tasks relevant to space engineering. In this work, only those informative features are required to be generated in order to perform classification. This minimizes feature measurement noise and the computational complexity of both feature extraction and feature vector-based classification. Experimental results show that this application ensures rapid and accurate learning of classifiers. This is of great importance to on-board image classification in future Mars rover missions.

The rest of this paper is organized as follows. Section 2 introduces the Mars terrain images under investigation. Sections 3, 4 and 5 outline the key component techniques used in this work, including feature extraction, (fuzzy-rough) feature selection and feature pattern classification. Section 6 shows the experimental results, supported by comparative studies. The paper is concluded in Section 7, where prospects for further research are discussed.

2 *McMurdo* Panorama Image

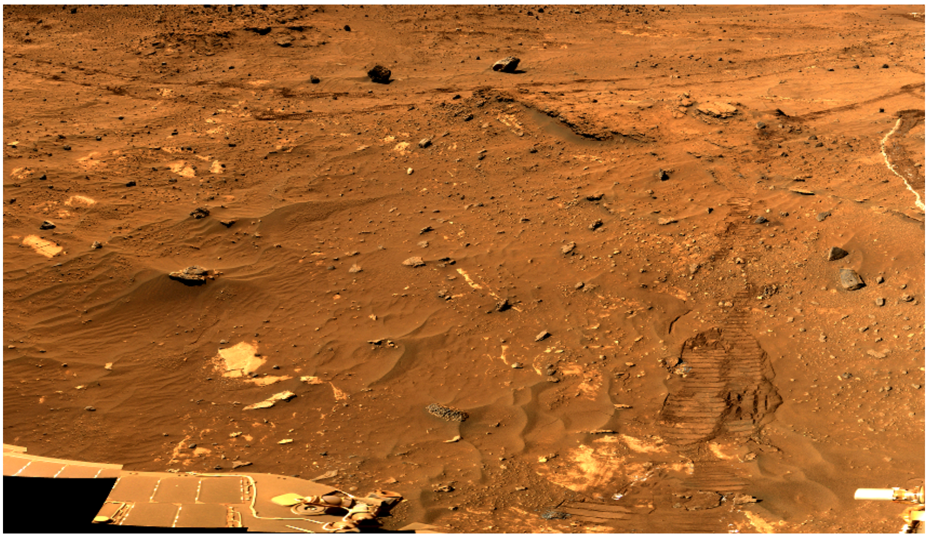
Although the approach taken in this research is general, the present work concentrates on the classification of the 360-degree view *McMurdo* panorama image. This (composed) image is obtained from the panoramic camera on NASA's Mars Exploration Rover Spirit and presented in approximately true color [12], consisting of 1,449 Pancam images and representing a raw data volume of nearly 500 megabytes. Such an image reveals a tremendous amount of detail in part of Spirit's surroundings, including many dark, porous-textured volcanic, brighter and smoother-looking rocks, sand ripple, and gravel (mixture of small stones and sand).

Figure 1 shows the most part of the original *McMurdo* image (of a size 20480×4124), which is separated into three portions to ensure quality of display. This image, excluding the areas occupied by the instruments and black shadows, is used for the work here, involving five major image types (i.e. classes) which are of particular interest. These image types are: textured or smoothed dark rock (C1), orange colored bedding rock (C2), light gray rock (C3), sand (C4), and gravel (C5), which are illustrated in Fig. 2. The ultimate task of this research is to develop an image classifier that can detect and recognize these five types of regions within a given image.

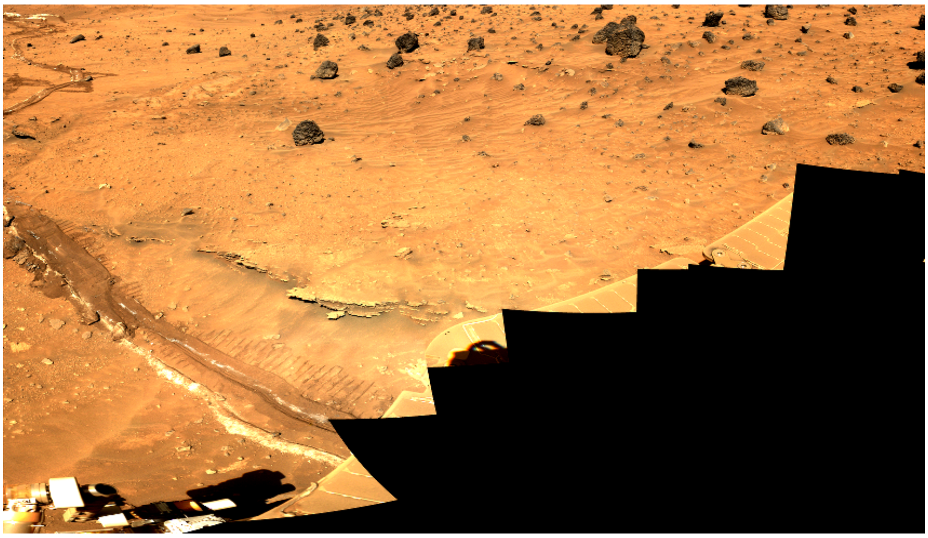
3 Feature Extraction

Feature extraction involves simplifying the amount of memory and computation power required to describe complex images accurately. A variety of techniques may be used to capture and represent the underlying characteristics of a given image [11, 23]. These include but are not limited to algorithms for:

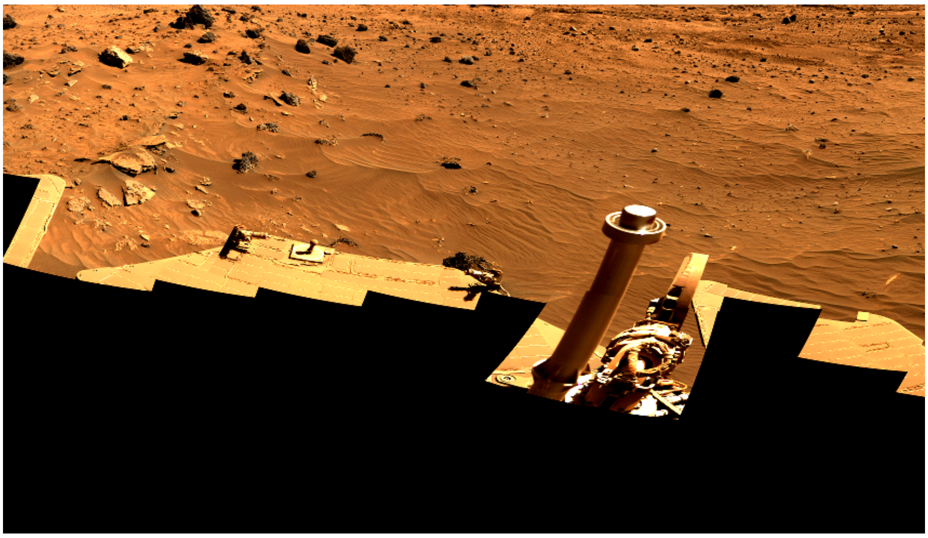
– Low-level detection of edge and corner which aims at identifying points at which the image brightness changes sharply, and that of blob and ridge which aims at finding regions in the image that are either brighter or darker than the surrounding [19];
– Curvature determination which aims at describing repeating patterns through cross-correlation of points or regions, and motion detection which aims at measuring change in speed of an object or objects in the field of view [8]; and
– Shape-based template matching which aims at deciding on which small parts of an image match a given template image [1], and Hough transform which aims at detecting positions of arbitrary shapes, most commonly lines, circles or ellipses via a voting procedure carried out in a certain parameter space [9].



(a) Portion 1



(b) Portion 2



(c) Portion 3

Fig. 1. Mars McMurdo panorama image.

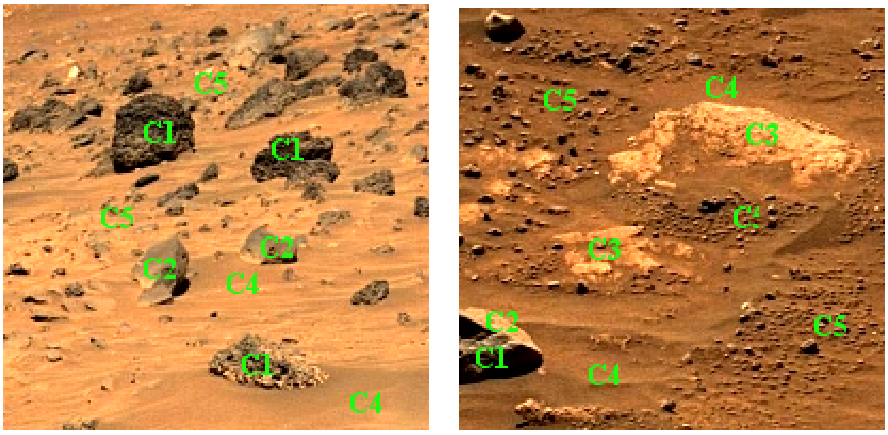


Fig. 2. Image classes (C1: rock1, C2: rock2, C3: rock3, C4: sand, C5: gravel).

In this work, low-level feature extraction approaches are employed. In particular, local grey level histograms and the first and second order color statistics are exploited to produce a feature vector for each individual pixel. This is due to the recognition that such features are effective in depicting the underlying image characteristics and are efficient to compute. Also, the resulting features are robust to image translation and rotation, thereby potentially suitable for classification of Mars images. Below is a brief outline of the techniques that are adopted for the implementation of feature extraction in this research.

3.1 Color Statistics-Based Features

Color images originally given in the RGB (Red, Green and Blue) space are first transformed to those in the HSV (Hue, Saturation and Value) color space [23]. These spaces are in bijection with one another, and the HSV space is widely used in the literature. Note that HSV is more frequently used because it separates the color components (H and S) from the luminance component (V) and is less sensitive to illumination changes [3]. Also, distances in the HSV space tend to reflect perceptual differences in color in a more consistent way than those measured in the RGB space.

By computing the first order (mean) and the second order (standard deviation, denoted by STD) color statistics with respect to each of the H, S and V channels, six features can be generated per pixel, from a certain neighborhood of that pixel. The size of such neighborhoods is pre-selected by trial and error (which trades off between the computational efficiency in measuring the features and the representative potential of the measured features). For presentational simplicity, the resulting features are hereafter denoted as Mean(X) and STD(X), X $\in$ {H, S, V}, representing the first and second order statistics per color channel respectively.

3.2 Local Histogram-Based Features

As the name indicates, such features are measured off the histograms computed from local regions of a given image. Although the method is very simple, it has shown very good results in practice, especially in performing tasks where feature extraction has to be as simple and as fast as possible [7] (which is indeed required for Mars missions where any possible efficiency gain is significant).

In the present context, a histogram is a summary graph showing a count of grey levels falling in a number of resolution ranges (called bins), within a predefined neighborhood. To extract histogram-based features, given color images are first transformed to grey-level (GL) images. For a certain pixel, a set of histogram features $H_i, i = 1, 2, ..., B$, are calculated (within the given neighborhood), with respect to a particular bin size B (i.e. number of bins). Thus, feature H_i represents the normalized frequency of the GL histogram in bin i. Here, for simplicity, individual bin widths are set equally, and the neighborhood size is set to the same as that used in the above color feature extraction.

Note that given certain image data, it is important to choose an appropriate bin size. If it is too small, a bar height at each bin suffers significant statistical fluctuation due to paucity of grey-levels in each bin. If however, the bin size is too large, a histogram cannot represent the shape of the underlying distribution because the resolution is not sufficient. To balance between these two problems, bin size B is empirically set to 16 in this work.

In addition to histogram features, two further GL statistic features are also generated, namely, the mean and STD. These are different from their color statistics-based counterparts and are denoted by Mean(GL) and STD(GL) for short.

4 Feature Selection

Feature selection (FS) [15, 20] addresses the problem of selecting amongst given features that are most informative. Fig. 3 shows the basic procedures involved in such a process. It can be seen from this figure that FS forms a particular approach to the reduction of the number of features under consideration. Importantly, unlike conventional dimensionality reduction or feature extraction methods, a feature selection algorithm preserves the original meaning of the features after reduction.

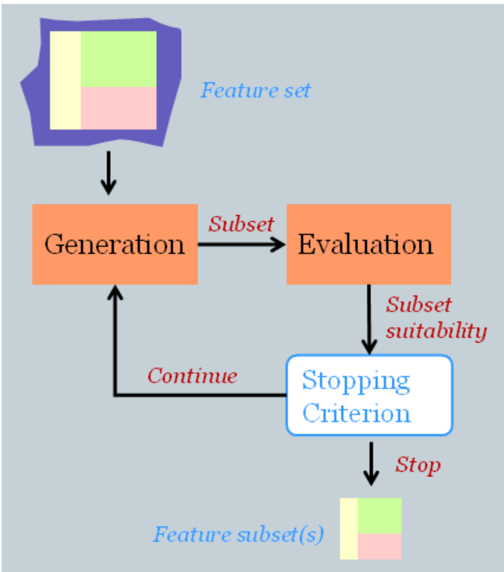


Fig. 3. Feature Selection Process

In general, there may be many features involved in Mars image analysis, and combinatorially large numbers of feature combinations to select from. It might be expected that the inclusion of an increasing number of features would increase the likelihood of including enough information to distinguish between classes. Unfortunately, this is not necessarily true if the size of the training dataset does not also increase rapidly with each additional feature included. A high-dimensional feature vector representation increases the chances that a classifier-learning algorithm finds spurious patterns that are not valid. More features may introduce more measurement noise and, hence, reduce model accuracy [13].

4.1 Fuzzy-Rough Feature Selection

Recently, there have been significant advances in developing methodologies that are capable of minimizing feature subsets in a noisy environment. In particular, a resounding amount of research utilizes fuzzy and rough sets [24, 25] (see [5, 15, 21, 22] for examples). Amongst them is the fuzzy-rough feature selection algorithm [14] that has been shown to be a highly useful technique by which discrete or real-valued noisy data (or a mixture of both) can be effectively reduced, without the need for user-supplied information. Inspired by this observation, FRFS is utilized in this work to maximize both classification efficiency and effectiveness, while ensuring that classification is carried out with (a subset of) original features only. The theoretical foundation of this feature selection method is outlined below.

Let U be the set of pixels within a given image, P be a subset of features, and D be the set of all possible image classes of interest. The concept of fuzzy-rough dependency measure, of D upon P (which FRFS is based on), is defined by [15]:

$$\gamma_P(D) = \frac{\sum_{x \in U} \mu_{POS_{R_P}(D)}(x)}{|U|} \tag{1}$$

where

$$\mu_{POS_{R_P}(D)}(x) = \sup_{X \in U/D} \mu_{\underline{R_P}X}(x) \tag{2}$$

$$\mu_{\underline{R_P}X}(x) = \inf_{y \in U} I(\mu_{R_P}(x, y), \mu_X(y)) \tag{3}$$

and U/D denotes the (equivalence class) partition of the image (i.e. pixel set) with respect to D, and I is a fuzzy implicator and T a t-norm [15]. R_P is a fuzzy similarity relation induced by the feature subset P:

$$\mu_{R_P}(x, y) = T_{A \in P}\{\mu_{R_{\{A\}}}(x, y)\} \tag{4}$$

That is, $\mu_{R_{\{A\}}}(x, y)$ is the degree to which pixels x and y are deemed similar with regard to feature A. It may be defined in many ways, but in this work, the following commonly used similarity relation [14] is adopted:

$$\mu_{R_{\{A\}}}(x, y) = 1 - \frac{|A(x) - A(y)|}{A_{max} - A_{min}} \tag{5}$$

where $A(x)$ and $A(y)$ stand for the value of feature $A \in P$ of pixel x and that of y, respectively, and A_{max} and A_{min} are the maximum and minimum values of feature A.

FRFS works by employing the above dependency measure to choose which features to add to the subset of the current best features through a greedy hill-climbing process. As such, the criterion that the algorithm adopts to select a subset of features is whether the set of the underlying true image classes may be maximised using those selected features. It terminates when the addition of any remaining feature does not increase the dependency. Thus, all features selected are individually significant, although significance embedded in any correlated features may not be necessarily captured. This implies that the feature subsets returned may not be globally minimal. Yet, this general problem is due to the use of greedy hill-climbing search, not because of the utilisation of fuzzy-rough dependency measure.

There exist sophisticated approaches to identifying or approximating the absolute smallest subsets of features [5]. However, in general, searching for all minimal feature subsets is an NP-complete problem. In practice, it may suffice to generate only one such subset or even a superset of some of such subsets (if it is too time consuming to compute a global minimal). The present work follows this theme of consideration. Indeed, FRFS takes the same approach as the seminal QuickReduct algorithm [4] which is based on conventional rough set theory. Empirically (see experimental results later), the use of fuzzy-rough dependency measure in conjunction with greedy hill-climbing provides an effective means to find quality feature subsets.

Algorithm 1 outlines the fuzzy-rough feature selection method. What is returned by this algorithm is the subset of features selected from the full set of original features without altering the meaning and value of such selected features.

Algorithm 1 The Fuzzy-Rough Feature Selection (FRFS) Algorithm

FRFS(C, D): C – the set of all original features; D – the set of possible image classes.

$R \leftarrow \{\}$, $\gamma_{best} = 0$
do
    $T \leftarrow R$, $\gamma_{prev} \leftarrow \gamma_{best}$
    foreach $a \in (C - R)$
        if $\gamma_{R \cup \{a\}}(D) > \gamma_T(D)$
            $T \leftarrow R \cup \{a\}$, $\gamma_{best} \leftarrow \gamma_T D$
    $R \leftarrow T$
until $\gamma_{best} == \gamma_{prev}$
return R

5 Image Classifiers

Multi-layer perceptron (MLP) neural networks [26] and K-nearest neighbors (KNN) [6] are used here to accomplish image classification, by mapping input feature vectors onto the underlying image class labels. As indicated previously, these advanced techniques are used to minimize the mission risk that might incur if more research-oriented classification systems were employed (which often require further comprehensive experimental evaluation). For learning such classifiers, a set of training data is selected from the typical parts (see Fig. 2) of the *McMurdo* image, with each pixel represented by a feature vector which is manually assigned an underlying class label.

5.1 Multi-Layer Perceptrons

Multi-layer perceptrons are well-known mechanisms to implement pattern classifiers. The design of such an image classifier is straightforward: The number of nodes in its input layer is set to that of the cardinality of the feature subset returned by FRFS, and the number of nodes within its output layer is set to the number of image classes for the problem at hand. The internal structure of the classifier is designed to be flexible and may contain one or two hidden layers. (What actual number of internal layers and that of hidden nodes in each hidden layer would be better to use may be determined by experimental simulations given a fixed number of input features and training images.)

The training of such a classifier is essential to its runtime performance (done here by using the back-propagation algorithm [26]). For this, feature patterns that represent different regions of the *McMurdo* panorama image, coupled with their respective underlying image class indices, are selected as the training data. Note that all input features are normalized into the range of 0 to 1 to suit the use of multi-layer perceptrons.

5.2 K-Nearest Neighbors

K-nearest neighbors (KNN) algorithm is one of the simplest learning methods for building classifiers. To classify an unclassified feature vector X, KNN ranks the neighbors of X amongst a given set of N data (X_i, c_i), $i = 1, 2, ..., N$, and uses the class labels c_j ($j = 1, 2, ..., K$) of the K most similar neighbors to predict the class of the new vector X. In particular, the classes of these neighbors are weighted using the similarity between X and each of its neighbors, where similarity is typically measured by the Euclidean distance metric (though any other distance metric may also do). Given such similarity measures, X is assigned the class label with the greatest number of votes amongst the K nearest class labels.

Note that KNN does not rely on prior probabilities, and that it is computationally efficient if the dimensionality of the input features is not very large. If however, the dimensionality is high, each distance calculation may become quite expensive. This reinforces in general, the need for employing FRFS to perform feature selection, in order to reduce the computation cost.

6 Experimental Results

From the *McMurdo* image of Fig. 1, a set of 270 subdivided non-overlap images with a size of 512 $\times$ 512 each are used to perform this experiment. 816 pixel points are selected from 28 of these images. Each of the pixels is labeled with an identified class index (i.e. one of the five image types: rock1, rock2, rock3, sand and gravel) for training and verification. The rest of all these images are used as unseen data for classification. Each training pixel is represented by a vector of 24 features (i.e. 6 color statistics-based, 16 histogram-based, and 2 grey-level statistic features, see Section 3). Of course, the actual classification process only uses subsets of selected features (which are returned by running FRFS). The performance of each classifier is measured using classification accuracy, with ten-fold cross validation.

Table 1. Feature meaning and reference

Ref No.	Meaning	Ref No.	Meaning	Ref No.	Meaning
1	Mean(GL)	2	STD(GL)	3	Mean(H)
4	STD(H)	5	Mean(S)	6	STD(S)
7	Mean(V)	8	STD(V)	9-24	H_i

For easy cross-referencing, Table 1 lists the reference numbers of the original features that may be extracted, where $i = 1, 2, ..., 16$, with respect to their meaning (see Section 3). Note that in the following, for KNN classification, the results are first obtained with K set to 1, 3, 5, 8, and 10. Also, for multi-layer perceptron-based classifiers, only those of one hidden layer are considered here (to limit simulation cost) with the number of hidden nodes first set to 8, 12, 16, 20, or 24. Those classifiers which have the highest accuracy, with respect to a given feature pattern dimensionality and a certain number of nearest neighbors or hidden nodes, are then taken to run for performance comparison. Here, comparative studies involve: (1) evaluation of the implication and hence, quality of using FRFS-selected features as opposed to the use of all original features, and (2) examination of the class discrimination power of FRFS-selected features over that of using a popular alternative dimensionality reduction mechanism.

6.1 Comparison with the Use of all Original Features

This subsection shows that, at least, the use of a selected subset of features does not significantly reduce the classification accuracy as compared to the use of the full set of original features. For the current problem, FRFS returns 8 features, namely, STD(GL), Mean(H), STD(H), Mean(S), STD(S), Mean(V), H_4, H_{15} (i.e. features $2, 3, 4, 5, 6, 7, 12$ and 23 in Table 1), out of the original twenty-four. This implies that the use of FRFS leads to a reduction rate of two-third. Table 2 lists the correct classification rates produced by the MLP and KNN classifiers with 10-fold-cross-validation, where the number (N) of hidden nodes and that (K) of the nearest neighbors used by these MLP and KNN classifiers are also provided (in the first column).

Table 2. FRFS-selected vs. full set of original features

Classifier	Set	Dim.	Feature No	Rate
MLP(N=20)	FRFS	8	$2, 3, 4, 5, 6, 7, 12, 23$	94.0%
MLP(N=20)	Full	24	$1, 2, ..., 23, 24$	94.0%
KNN(K=8)	FRFS	9	$2, 3, 4, 5, 6, 7, 12, 23$	89.1%
KNN(K=5)	Full	24	$1, 2, ..., 23, 24$	89.2%

These results demonstrate that the classification accuracy of using the eight FRFS-selected features is the same as that of using the twenty-four original features for MLP classifiers (94.0%), and is very close to that for KNN classifiers (89.1% vs. 89.2%). This is indicative of the potential of FRFS in reducing not only redundant feature measurements but also the noise associated with such measurements. Clearly, the use of FRFS helps to improve both effectiveness and efficiency of the classification process. Note that although the number of original features is not large, for on-board Martian application, especially in relation to the task of classifying large-scale images, any reduction of feature measurements is of great practical significance.

6.2 Comparison with the Use of PCA-Returned Features

Principal component analysis (PCA) [6] is arguably one of the most popular methods for dimensionality reduction (although it was not initially designed to obtain discriminatory features). It is therefore adopted here as the benchmark for comparison. Fig. 4 shows the classification results of the KNN and MLP classifiers using a different number of principal features. For easy comparison, the results of the KNN and MLP that use 8 FRFS-selected features are also included in the figure, which are represented by $*$ and $\circ$, respectively.

These results show that the MLP classifier which uses FRFS-selected features has a substantially higher classification accuracy amongst all those classifiers using a subset of features of the same dimensionality (i.e. 8). This is achieved via a considerably simpler computation, due to the substantial reduction of the complexity in input feature vectors. The results also show the cases where PCA-aided (MLP or KNN) classifiers each employ a feature subset of a different dimensionality. However, these classifiers still generally underperform than their FRFS-aided counterparts, whether they are implemented using MLP or KNN. This situation only changes when almost the full set of PCA-returned features is used where the MLP classifiers may perform similarly or slightly better (if 20

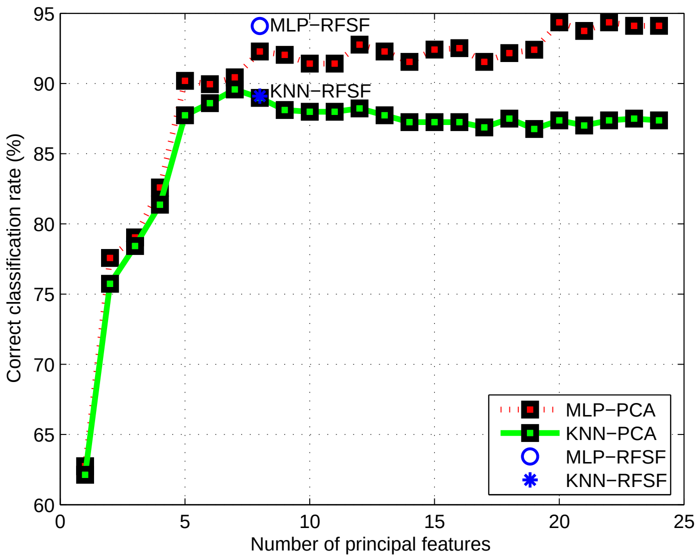


Fig. 4. Performance of KNN and MLP vs. the number of principal components.

or 22 principal components are used). Yet, this is at the expense of requiring many more feature measurements and much more complex classifier structures.

Additionally, it is worth noting that unlike FRFS-selected features, those returned by PCA lose the underlying meaning of the original features. They are the linear combinations of the original instead. This may cause any subsequent analysis and explanation of the image classification results more difficult to the user [15].

6.3 Classified and Segmented Images

The ultimate task of this research is to classify Mars panoramic camera images and to detect different regions or objects (e.g. rocks) in such images. Based on the above results, the MLP which employs the 8 FRFS-selected features, and which was trained by the given 816 labeled feature patterns, is taken to accomplish this task: the classification of the entire image of Fig. 1 (excluding certain regions as indicated previously).

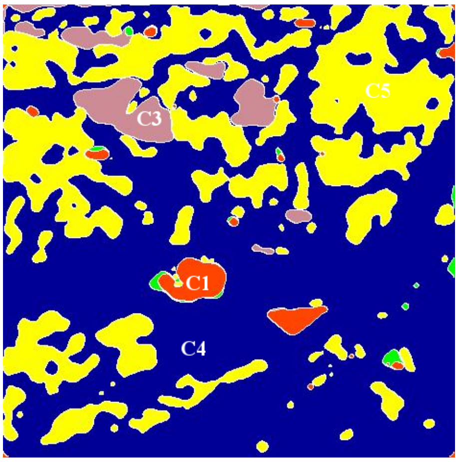
As an illustration, four classified images are shown in Fig. 5, which are numbered by (a), (c), (e) and (g) respectively, and presented on the left side of this figure, where five different colors represent the five image types (namely, rock1, rock2, rock3, sand and gravel). From this, boundaries between different class regions can be identified and marked with white lines, resulting in the segmented images also given in Fig. 5, which are numbered by (b), (d), (f) and (h) correspondingly, and presented on the right.

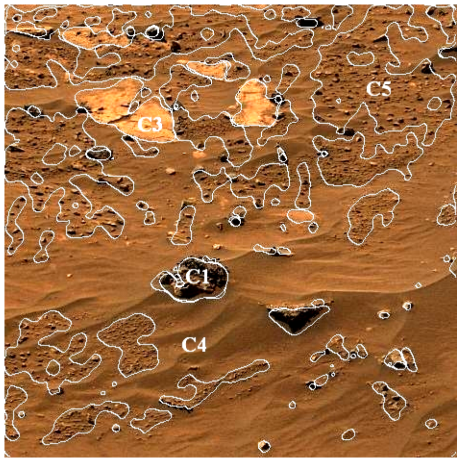
From these classified images, it can be seen that the five image types vary in terms of their size, rotation, color, contrast, shape, and texture. For human eyes it can be difficult to identify boundaries between certain image regions, such as those between sand and gravel, and those between rock2 and sand. However, the classifier is able to perform under such circumstances, showing its robustness to image variations. This indicates that the small subset of features selected by FRFS indeed convey the most useful information of the original. Note that classification errors mainly occur within regions representing sand and gravel. This may be expected since gravel is itself a mixture of sand and small stones.

7 Conclusion

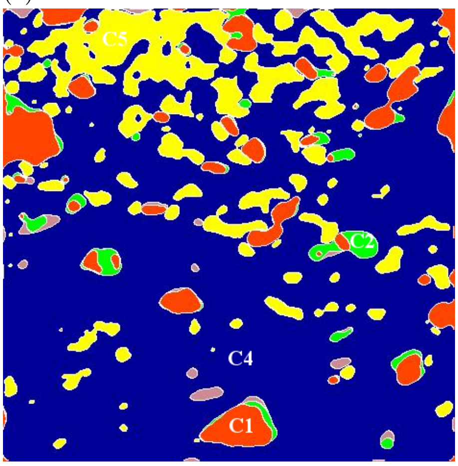
This paper has presented a study on Mars terrain image classification. In particular, advanced fuzzy-rough feature selection techniques have been adopted to help solving such important, and practically challenging, problems in space engineering. Although the real-world images encountered are large-scale and complex, the resulting dimensionality of selected feature vectors is manageable. Conventional classifiers such as MLP and KNN that are built using such selected features generally outperform those using more features or an equal number of features obtained by classical approaches represented by PCA. This is confirmed by systematic experimental investigations. The work helps to identify mature mechanisms that can accomplish Martian image classification effectively and efficiently. This is of particular significance for modeling and analysis of real images on board in future Mars rover missions, as only advanced and proven efficient techniques will be chosen to conduct real mission tasks.

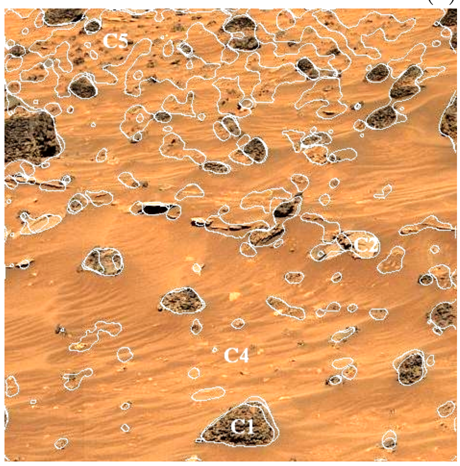
Whilst this work is very promising, active research remains to enhance the performance of FRFS-assisted classifiers. For instance, the influence of parameter set-up for feature extraction, e.g. the neighborhood size and bin size, requires further investigation. A more systematic selection of the number of bins used is indeed a general issue for any histogram-based feature extraction approaches [30]. Also, it would be very interesting to compare the present work with the use of alternative feature selection methods (e.g. information gain-based [27]) or classification techniques (e.g. rough-neural hybridization [10]). Finally, currently, histogram-based features are produced using grey-level images (which have been transformed from the original color images). Work on direct use of color images to generate color-channel-based histogram features may help to capture more essential pattern discriminatory power though it will also introduce extra computation cost. An investigation into how to appropriately balance between effectiveness and efficiency is on-going. Again, the employment of FRFS may well offer significant assistance in choosing, and hence utilizing, just the most informative color histogram features.



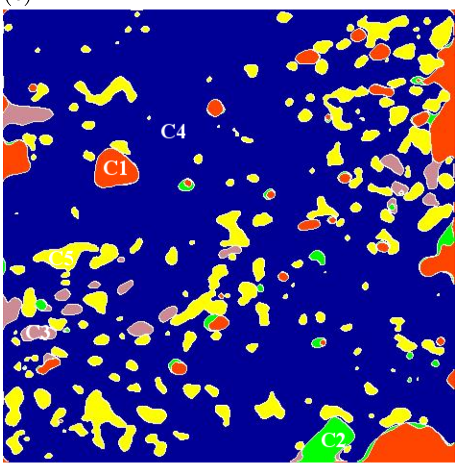


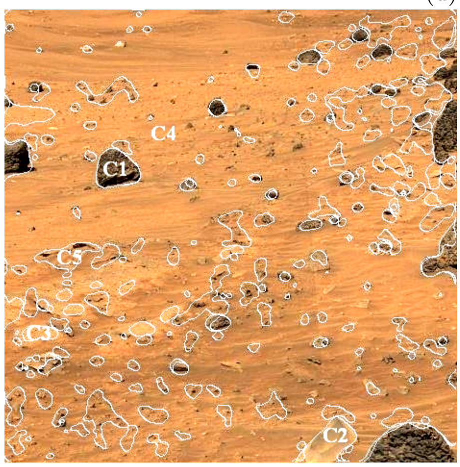
(a) (b)



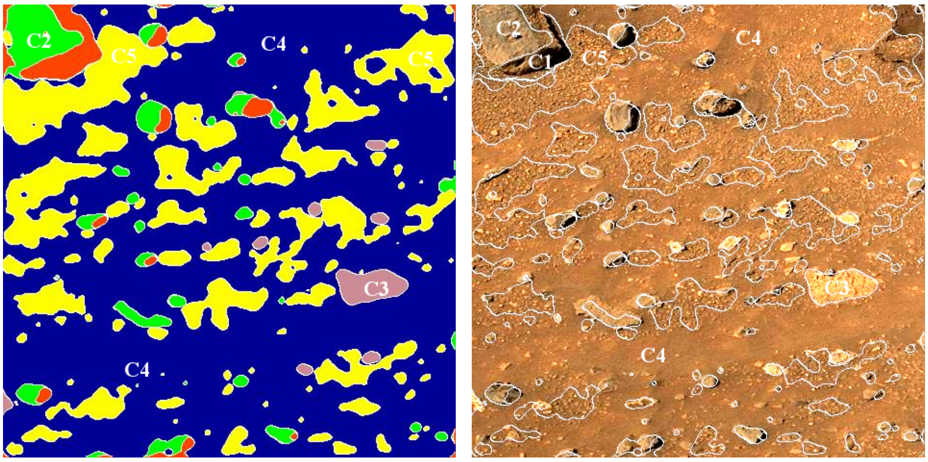


(c) (d)





(e) (f)



(g) (h)

Fig. 5. Classified (left) and segmented (right) images.

Acknowledgements

This work was partly supported by a UK Royal Academy of Engineering-sponsored Daphne Jackson Fellowship. The authors are grateful to reviewers for their comments which have helped to improve this work.

References

1. Brunelli R.: Template Matching Techniques in Computer Vision: Theory and Practice, Wiley (2009)
2. Castano R. *et al.*: Current results from a rover science data analysis system. Proc. of IEEE Aerospace Conf. (2006)
3. Chapelle O., Haffner P., Vapnik V. N.: Support vector machines for histogram-based image classification. IEEE Transactions on Neural Networks. 10(5):1055-1064 (1999)
4. Chouchoulas A., Shen Q.: Rough set-aided keyword reduction for text categorisation. Applied Artificial Intelligence 15(9):843-873 (2001)
5. Cornelis C., Jensen R., Martin G., Slezak D.: Attribute selection with fuzzy decision reducts. Information Sciences. 180(2):209-224 (2010)
6. Duda R. O., Hart P. E., Stork D. G.: Pattern classification. (2nd edition). Wiley & Sons, New York (2001)
7. Dunlop H., Thompson D., Wettergreen D.: Multi-scale features for detection and segmentation of rocks in Mars images. IEEE Int. Conf. CVPR. 1-7 (2007)
8. Dunn P. F.: Measurement and Data Analysis for Engineering and Science. New York: McGrawHill (2005)
9. Hart P. E.: How the Hough Transform was Invented. IEEE Signal Processing Magazine. 26(6):18-22 (2009)

10. Hassanien A. E., Slezak D.: Rough neural intelligent approach for image classification: A case of patients with suspected breast cancer. Int. J. Hybrid Intell. Syst. 3(4):205-218 (2006)
11. Huang K., Aviyente S.: Wavelet feature selection for image classification. IEEE Trans. Image Proc. 17, 1709-1720 (2008)
12. http://marswatch.astro.cornell.edu/pancam_instrument/mcmurdo_v2.html
13. Jensen, R., Shen, Q.: Are more features better? IEEE Transactions on Fuzzy Systems. 17(6):1456-1458 (2009)
14. Jensen R., Shen Q.: New approaches to fuzzy-rough feature selection. IEEE Trans. Fuzzy Syst., 17(4):824-838 (2009)
15. Jensen R., Shen Q.: Computational intelligence and feature selection: rough and fuzzy approaches. IEEE Press and Wiley (2008)
16. Kachanubal T., Udomhunsakul S.: Rock textures classification based on textural and spectral features. Proc. of World Academy of Science, Eng. and Tech. 29, 110-116 (2008)
17. Kim W. S., Steele R. D., Ansar A. I., Al K., Nesnas I. Rover-Based visual target tracking validation and mission infusion. AIAA Space. 6717-6735 (2005)
18. Lepisto L., Kunttu I., Visa A.: Rock image classification based on k-nearest neighbour voting. Vis. Im. and Sig. Proc., IEE Proc. 153(4), 475-482 (2006)
19. Lindeberg T.: Edge detection and ridge detection with automatic scale selection. International Journal of Computer Vision. 30(2):117-154 (1998).
20. Liu H., Motoda H. Feature Selection for Knowledge Discovery and Data Mining. Kluwer Academic Publishers (1998)
21. Mac Parthalain, N., Shen, Q.: Exploring the boundary region of tolerance rough sets for feature selection. Pattern Recognition, 42(5):655-667 (2009)
22. Mac Parthalain, N., Shen, Q., Jensen, R.: A distance measure approach to exploring the rough set boundary region for attribute reduction. IEEE Transactions on Knowledge and Data Engineering. 22(3):306-317 (2010)
23. Martin D. R., Fowlkes C. C., Malik J.: Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Trans. Patt. Anal. and Mach. Inte. 26:530-549 (2004).
24. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishing (1991)
25. Pawlak, Z., Skowron, A.: Rudiments of rough sets. Information Sciences. 177(1):3-27 (2007)
26. Rumelhart D., Hinton E., Williams R.: Learning internal representations by error propagating. In: Rumelhart D., McClell J. and (Eds.): Parallel Distributed Processing. MIT Press (1986)
27. Shang C., Barnes D.: Combining support vector machines and information gain ranking for classification of Mars McMurdo panorama images, Proc. of IEEE International Conference on Image Processing, (2010)
28. Shang C., Barnes D., Shen Q.: Taking fuzzy-rough application to Mars: Fuzzy-rough feature selection for Mars terrain image classification. Proc. of 12th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, 209-216 (2009)
29. Shang C., Barnes D., Shen Q.: Effective feature selection for Mars McMurdo image classification. Proc. of 9th International Conference on Intelligent Systems Design and Applications, 1419-1424 (2009)
30. Shimazaki H., Shinomoto S.: A method for selecting the bin size of a time histogram Neural Computation. 19(6):1503-1527 (2007)

31. Thompson D. R., Castano R.: Performance comparison of rock detection algorithms for autonomous planetary geology. Proc. of IEEE Aerospace Conf. paper no. 1251 (2007)
32. Woods M., Shaw A., Barnes D., Price D., Long D., Pullan D.: Autonomous science for an ExoMars Rover-like mission, Journal of Field Robotics, 26(4): 358-390 (2009)